CIFellows 2020-2021

Computing Innovation Fellows

Kevin Bello

University of Chicago & Carnegie Mellon University

Faster, Better, and Simpler: Learning DAGs via a log-determinant acyclicity characterization. (joint with: Bryon Aragam and Pradeep Ravikumar)

1.Motivation:

Bayesian networks are a popular graphical model with applications in biology, machine learning, causal inference, etc.

2. Approach:

We follow the score-based approach, i.e., assign a score to each DAG. Then, the **goal** is to select the DAG with **lowest score**.

3. Challenge:

The space of DAGs is **discrete** with **cardinality super-exponential** on the number of variables.

4. Previous work:

Use the following differentiable function

$h_{\exp}(W) = \operatorname{Tr}\left(e^{W \circ W}\right)$

5.Proposed method:

Use instead the following differentiable function

$h(W) = -\log\det(I - W \circ W)$

You can learn directed acyclic graphs (DAGs; a.k.a. Bayesian networks) from data via continuous optimization.

$$\min_{W \in \mathbb{R}^{d \times d}} F(W)$$

subject to $G(W) \in$ **DAGs**

Discrete space

$$\min_{W \in \mathbb{R}^{d \times d}} F(W)$$

subject to $h(W) = 0$

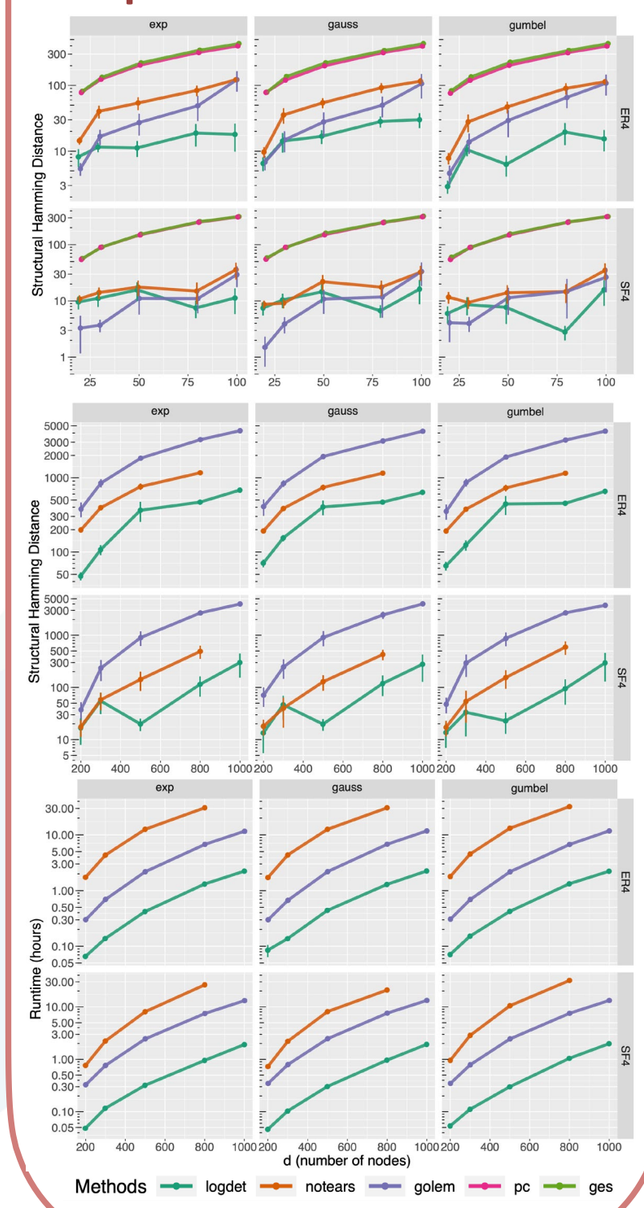
Continuous function!

This work is about learning DAGs faster using a simpler framework and obtaining better accuracy.

6. Why?

- $h(W)$ does not diminish cycles of any length.
- $h(W)$ has a stronger gradient than $h_{\exp}(W)$.
- $h(W)$ has a simpler gradient $\nabla h(W) = 2(I - W \circ W)^{-\top} \circ W$.

7. Experiments



CRA Computing Research Association

CCC Computing Community Consortium Catalyst

NSF